Web-based Supplementary Materials for "Single-index Varying Coefficient Model for Functional Responses"

Xinchao Luo[1,2], Lixing Zhu[3], and Hongtu Zhu[2*]

[1]School of Finance and Statistics, East China Normal University, Shanghai, China

[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

*hzhu@bios.unc.edu

[3]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

Web Appendix A: Theoretical Results

Assumptions

Throughout the paper, we need the following assumptions to facilitate the technical details, including traditional conditions for asymptotic theory, although they may not be the weakest conditions.

Assumption (C1). $\sup_{s_m} E[|\varepsilon_i(s_m)|^\kappa] < \infty$ for some $\kappa > 4$ and all grid points s_m.

Assumption (C2). The data $\{X_i, Y_i(s_m) : i = 1, \cdots, n, s_m \in \mathcal{S}_M\}$ are independently and identically distributed.

Assumption (C3). For all $s \in \mathcal{S}$, $\beta_0(s) \in \Theta$ is a unique point satisfying $E\{S_{\text{eff}}(\beta(s); X_i, Y_i(s))\} = \mathbf{0}$, where the expectation is taken with respect to the true distribution of $Y(s)$ given X. Moreover, $E\{\partial S_{\text{eff}}(\beta(s); X_i, Y_i(s))/\partial\beta(s)\}$ is nonsingular.

Assumption (C4). Θ is a compact set. For all $s \in \mathcal{S}$ and $\beta(s) \in \mathcal{B}$, $S_{\text{eff}}(\beta(s); X_i, Y_i(s))$ is twice continuously differentiable on Θ. For all $j, k = 1, \cdots, p$, $E[|\partial_j S_{\text{eff}}(Y(s), X^T\beta(s), f_2)|] \leqslant \infty$, $|S_{\text{eff}}(\beta(s); X_i, Y_i(s))|$, $|\partial_j S_{\text{eff}}(\beta(s); X_i, Y_i(s))|$ and $|\partial_j \partial_k S_{\text{eff}}(\beta(s); X_i, Y_i(s))|$ are dominated by an integral function $G(Y(s), X)$ such that $E[\sup_{s\in\mathcal{S}} |G(Y(s), X)|^r] < \infty$ for a $r \geqslant 1$, where $\partial_j = \partial/\partial\beta_j(s)$, in which $\beta_j(s)$ is the j-th component of $\beta(s)$.

Assumption (C5). Each component of $\{\eta(s), s \in \mathcal{S}\}$, $\{\eta(s)^2, s \in \mathcal{S}\}$, and $\{S_{\text{eff}}(\beta(s); X_i, Y_i(s)), s \in \mathcal{S}\}$ are P-Donsker classes.

Assumption (C6). The grid points s_m are randomly generated from a density function $\pi(s)$. Moreover, $\pi(s) > 0$ for all $s \in \mathcal{S}$, and $\pi(s)$ has continuous twice derivative with bounded support [0,1].

Assumption (C7). The kernel function K is a continuous symmetrical bounded density function with support $[-1, 1]$ and satisfies that

$$\int_{-1}^{1} K(\mu)d\mu = 1, \quad \int_{-1}^{1} \mu K(\mu)d\mu = 0, \mu_2(K) = \int_{-1}^{1} \mu^2 K(\mu)d\mu < \infty, \quad v_0(K) = \int K^2(\mu)d\mu < \infty.$$

Assumption (C8). For all $s \in \mathcal{S}$, $\beta(s)$ and $\eta(s)$ have finite continuous twice derivatives. Moreover, $E[\sup_{s\in\mathcal{S}} |\eta(s)|^{r_1}] < \infty$ and $E\{\sup_{s\in\mathcal{S}}[|\dot{\eta}(s)| + |\ddot{\eta}(s)|]^{r_2}\} < \infty$ for some $r_1, r_2 \in (2, \infty)$.

Assumption (C9). Both M and n tend to infinity, $Mh \to \infty$, $h^{-1}|\log h|^{1-2/q_1} \leqslant M^{1-2/q_1}$ for $q_1 \in (2, 4)$, $nMh_1^4 \to \infty$, $nMh_1^5/\log(nM) < \infty$, $h_2^{-4}(\log n/n)^{1-2q_2} = o(1)$ for $q_2 \in (2, \infty)$, $h_2 = o(1)$, and $Mh_2 \to \infty$.

Assumption (C10). The bandwidths h_x and h_y satisfy $nh_x^4 h_y^4 \to 0$, $nh_x^4 h_y^2 \to 0$ and $nh_x h_y \to \infty$.

Assumption (C11). The density functions of $X_i^T\beta(\cdot)$ is bound away from 0 and ∞ on their support, $E\{X|X_i^T\beta(\cdot)\}$ and $g(X_i^T\beta(\cdot))$ have locally Lipschitz continuous derivatives.

Remark. Assumption (C1) requires the uniform bound on the high-order moment of $\varepsilon_{ij}(s_m)$ for all grid points s_m. Assumption (C2) is a relatively weak condition on the covariate vectors and their identical distribution are not essential. For each $s \in \mathcal{S}$, Assumptions (C3)-(C4) are generalizations of the standard conditions for ensuring asymptotic properties (e.g., consistency and asymptotic normality) of Z-estimators (Van der Vaart, 1998), where Assumption (C3) is an identifiable and nonsingular condition, and Assumption (C4) is a uniform smoothness and integration condition. Particularly, Assumption (C4) ensures that $S_{\text{eff}}(Y(s), X^T\beta(s), f)$ is uniformly integrable for all $s \in \mathcal{S}$. Assumption (C5) follows Zhu et al. (2012) to avoid smoothness conditions on the sample path, which are commonly assumed for simultaneous inference. Assumption (C6) is a weak condition on the random grid points (Zhu et al., 2012). In many neuroimaging applications, M is often much larger than n and for such large M, a regular grid of voxels is fairly well approximated by voxels generated by a uniform distribution in a compact subset of Euclidean space. For notational simplicity, we only state the theoretical results for the random grid points throughout the paper. Assumption (C7) is commonly assumed for kernel smooth methods. Assumption (C8) is the smoothness condition of $\beta(s)$ and $\eta(s)$. Assumptions (C9) on bandwidths are similar to the conditions used in Zhu et al. (2012). Assumptions (C10)-(C11) are assumed by Ma and Zhu (2014) in order to establish the asymptotic properties of $\hat{\beta}(\cdot)$ at each grid point.

Asymptotic Properties

THEOREM 1: *We have the following results.*

(i) Suppose that $\beta(s) = \beta_0$ does not vary across $s \in \mathcal{S}$. The optimal $\mathbf{w}_*$ is given by

$$\mathbf{w}_* = \Sigma_{\epsilon*,M}^{-1}\mathbf{1}_M/||\Sigma_{\epsilon*,M}^{-1}\mathbf{1}_M||_2, \tag{1}$$

where $||\cdot||_2$ is the Euclidean norm of a vector, $\Sigma_{\epsilon*,M} = \Sigma_{\eta*,M} + \Lambda_{\epsilon*,M}$ is an $M \times M$ matrix, and

$\mathbf{1}_M$ *is an* $M \times 1$ *vector of ones. Thus, the optimal* $D(\mathbf{w})$ *is given by* $D(\mathbf{w}_*) = (\mathbf{1}_M^T \Sigma_{\epsilon*,M}^{-1} \mathbf{1}_M)^{-1}$ *and is independent of s. The* $\Sigma_{\epsilon*,M}^{-1}\mathbf{1}_M$ *can be written as*

$$\{\mathbf{I}_M - \Lambda_{\epsilon*,M}^{-1/2} \sum_{m=1}^{M} \frac{\lambda_{m*,M}}{1+\lambda_{m*,M}} \psi_{m*,M}\psi_{m*,M}^T \Lambda_{\epsilon*,M}^{1/2}\}\Lambda_{\epsilon*,M}^{-1}\mathbf{1}_M. \tag{2}$$

(ii) Suppose that $\beta(s)$ *may vary across* $s \in \mathcal{S}$*. Under Assumptions (C6) and (C7), if* $w(s_m, s) = K_h(s_m - s)$*,* $h \to 0$*, and* $Mh \to \infty$*, then* $D(\mathbf{w}(s))$ *can be approximated by* $\Sigma_\eta(s,s)$*.*

THEOREM 2: *Under Assumptions (C1)-(C11), as* $n, M \to \infty$*, we have the following results.*

(i)

$$\sqrt{n}\left(\hat{\beta}(s) - \beta_0(s) - 0.5h^2 A_n(s)^{-1}\{\frac{1}{n}\sum_{i=1}^{n} B_i(s)[\ddot{g}(X_i^T\beta(s)) + 2\dot{g}(X_i^T\beta(s))\dot{\pi}(s)/\pi(s)]\mu_2(K)\}\right)$$

converges weakly to a Gaussian process with mean zero and covariance function, which is the limiting function of $A_n(s)^{-1}\left[n^{-1}\sum_{i=1}^n B_i(s)\Sigma_\eta(s,t)B_i(t)^T\right]A_n(t)^{-1}$*, where* $A_n(\cdot)$ *and* $B_i(\cdot)$ *are defined in Web Appendix C.*

(ii) $\sqrt{n}[\hat{g}(X^T\hat{\beta}(s)) - g(X^T\beta(s)) - 0.5h_1^2\mu_2(K)\ddot{g}(X^T\beta(s))]$ *converges weakly to a Gaussian process with mean zero and covariance function* $\Sigma_\eta(s,t)$*.*

We next study the asymptotic bias and covariance of $\hat{\eta}_i(s)$ as follows. We distinguish between two cases. The first one is to condition on the design points in $\mathcal{S}$, X and η. The other is to condition on the design points in $\mathcal{S}$ and X. The two cases may both be of interest for practitioners. We define $K^{\#}((s-t)/h) = \int K(u)K(u + (s-t)/h)du$.

THEOREM 3: *Under Assumptions (C1)-(C11), the following results hold for all* $s \in \mathcal{S}$*.*

(i) Conditioning on $(\mathcal{S}, X, \eta)$*, we have*

$$\begin{aligned} &\text{Bias}[\hat{\eta}_i(s)|\mathcal{S}, X, \eta] \\ =&0.5\mu_2(K)[\ddot{\eta}_i(s)h_2^2 + \ddot{g}(X_i^T\beta(s_m))h_1^2][1 + o_p(1)] + O_p(n^{-1/2}), \\ &\text{Cov}[\hat{\eta}_i(s), \hat{\eta}_i(t)|\mathcal{S}, X, \eta] \\ =&K^{\#}((s-t)/h_2)\sigma_\varepsilon(s)\pi(t)^{-1}(Mh_2)^{-1}O_p(1) - (nMh_1)^{-1}O_p(1). \end{aligned}$$

(ii) The asymptotic bias and covariance of $\hat{\eta}_i(s)$ conditional on $\mathcal{S}$ and X are given by

$$\mathrm{Bias}[\hat{\eta}_i(s)|\mathcal{S},X]=0.5h_1^2\mu_2(K)\ddot{g}(X_i^T\beta(s_m))[1+o_p(1)],$$

$$\mathrm{Cov}(\hat{\eta}_i(s)-\eta_i(s),\hat{\eta}_i(t)-\eta_i(t)|\mathcal{S},X)$$

$$=[1+o_p(1)][0.25\mu_2(K)^2h_2^4\Sigma_\eta^{(2,2)}(s,t)+K^{\#}((s-t)/h_2)\pi(t)^{-1}(Mh_2)^{-1}O_p(1)+n^{-1}\Sigma_\eta(s,t)].$$

(iii) The mean integrated squared error (MISE) of all $\hat{\eta}_i(s)$ is given by

$$\begin{aligned}&n^{-1}\sum_{i=1}^n\int_0^1E\{[\hat{\eta}_i(s)-\eta_i(s)]^2|\mathcal{S}\}\pi(s)ds\\ =&[1+o_p(1)]\times\{O((Mh_2)^{-1})+O_p(h_1^4)+n^{-1}\int_0^1\Sigma_\eta(s,s)\pi(s)ds\\ &+0.25\mu_2(K)^2\int_0^1\Sigma_\eta^{(2,2)}(s,s)h_2^4\pi(s)ds\}.\end{aligned}\tag{3}$$

(iv) The optimal bandwidth for minimizing MISE (3) is given by

$$\hat{h}_2=O(M^{-1/5}).\tag{4}$$

(v) The first order local polynomial kernel reconstructions $\hat{\eta}_i(s)$ using $\hat{h}_2$ in (4) satisfy

$$\sup_{s\in\mathcal{S}}|\hat{\eta}_i(s)-\eta_i(s)|=O_p(|\log(M)/5|^{1/2}M^{-2/5}+h_1^2+n^{-1/2})$$

for $i=1,\cdots,n$.

Remark. *Theorem 3 may be used to study the statistical property of the entity-specific effects in the functional regression analysis. Because the estimation of $\eta_i(s)$ succeeds the estimation of the regression coefficients, the MISE of $\hat{\eta}_i(s)$ appears to be a function of the two bandwidths h_1 and h_2. If the optimal bandwidth in Theorem 3 (iv) is used, the resulting MISE can achieve the order of $M^{-4/5}+h_1^4+n^{-1}$.*

The next theorem provides the asymptotic properties of the estimated covariance matrix and its spectrum decomposition.

THEOREM 4: *(i) Under Assumptions (C1)-(C11), it follows that*

$$\sup_{(s,t)\in\mathcal{S}^2}|\hat{\Sigma}_\eta(s,t)-\Sigma_\eta(s,t)|=O_p((Mh_2)^{-1}+h_1^2+h_2^2+(\log n/n)^{1/2}).$$

(ii) Under Assumptions (C1)-(C11), if the optimal bandwidths are used to construct $\hat{g}(x^T\hat{\beta}(s))$ *and* $\hat{\eta}_i(s)$*, then for* $1 \leqslant k \leqslant k_0$*, we have the following results:*

(a) $\int_0^1 [\hat{\psi}_k(s) - \psi_k(s)]^2 ds = O_p((Mh_2)^{-1} + h_1^2 + h_2^2 + (\log n/n)^{1/2})$*;*

(b) $|\hat{\lambda}_k - \lambda_k| = O_p((Mh_2)^{-1} + h_1^2 + h_2^2 + (\log n/n)^{1/2})$.

Remark. *A number of results concerning convergence rates of functional principal components analysis are available in the literature. Theorem 4 incorporates the recent development in Hall et al. (2006) and Zhu et al. (2012). The almost sure convergence result in Li and Hsing (2010); Li et al. (2010) seems to be slightly stronger since the authors consider a different approach to the estimation of* $\Sigma(s,t)$ *and their approach may not ensure* $\hat{\Sigma}(s,t)$ *to be positive semi-definite. Our results may be more appealing to practitioners, especially to imaging data analysts where acquiring a proper covariance estimation is key to the further inference.*

Some lemmas and their proofs

Introduce some notations first. Denote

$$\hat{\mathbb{M}}_{nM}(\beta(s)) = -\|\widehat{S}_{nM}(\hat{\beta}(s);w)\|,$$

$$\mathbb{M}(\beta(s)) = -\|E(S_{\text{eff}}(\beta(s);X_i,Y_i(s)))\|.$$

Denote $K_0(s,h) = \int K_h(t-s)\pi(t)dt$,

$$\Delta(s,h) = \sum_{m=1}^{M} \tilde{K}_h(s_m - s)\eta(s_m) - \frac{1}{K_0(s,h)}\int K_h(t-s)\eta(t)\pi(t)d(t).$$

LEMMA 1: *(Lemma 2 in Zhu et al. (2012)) Under Assumptions (C1), (C6), (C7) and (C9), for any* $r \geqslant 0$

$$\sup_{s\in\mathcal{S}}\left|\int K_h(t-s)\frac{(t-s)^r}{h^r}d[\Pi_M(t) - \Pi(t)]\right| = O_p((Mh)^{-1/2}),$$

$$\sup_{s\in\mathcal{S}}\left|\int K_h(t-s)\frac{(t-s)^r}{h^r}\varepsilon_{ij}(t)d\Pi_M(t)\right| = O_p((Mh)^{-1/2}\sqrt{|\log h|}),$$

where $\Pi_M(\cdot)$ *is the sampling distribution function based on* $\mathcal{S}_M = \{s_1,\cdots,s_M\}$, $\Pi(\cdot)$ *is the distribution function of* s_m.

LEMMA 2: *Under Assumptions (C1)-(C4) and (C6)-(C11),*

$$\sup_{s\in\mathcal{S}} d(\hat{\beta}(s), \beta_0(s)) \xrightarrow{P} 0.$$

Proof of Lemma 2: The proof consists of four steps.

Step 1: To prove $\hat{\beta}(s) \in \Theta$ with a probability going to one. Obviously $\hat{\beta}(s) \in [-1,1]^p$, next we show its continuity. Following the argument of Ma and Zhu (2014) we can obtain the pointwise consistency that for any $s \in \mathcal{S}$, $\hat{\beta}(s) \to \beta_0(s)$. Given two estimators $\hat{\beta}(s_1)$ and $\hat{\beta}(s_2)$ satisfying $|s_1 - s_2| < \delta$ for some $\delta \to 0$, so that $\|\beta_0(s_1) - \beta_0(s_2)\| \to 0$ under assumption (C8). On the other hand, since $\hat{\beta}(s_1) \to \beta_0(s_1)$ and $\hat{\beta}(s_2) \to \beta_0(s_2)$ in probability, we have $\|\hat{\beta}(s_1) - \hat{\beta}(s_2)\| \to 0$ by the triangle inequality in probability.

Step 2: To show the uniform convergence of $\hat{\mathbb{M}}_{nM}(\cdot)$ in probability, that is

$$\sup_{\beta(s)\in\Theta} |\hat{\mathbb{M}}_{nM}(\beta(s)) - \mathbb{M}(\beta(s))| \xrightarrow{P} 0. \tag{5}$$

Define $\psi(x) = x^{2p}$, by the mean value theorem we know that under Assumption (C4), $\hat{\mathbb{M}}_{nM}(\cdot)$ satisfies

$$\begin{aligned}
&\|\hat{\mathbb{M}}_{nM}(\beta_1(s)) - \hat{\mathbb{M}}_{nM}(\beta_2(s))\|_\psi \\
&= \|\hat{\mathbb{M}}_{nM}(\beta_1(s)) - \hat{\mathbb{M}}_{nM}(\beta_2(s))\|_{2p} \\
&\leqslant \mathrm{E}|\hat{\mathbb{M}}_{nM}(\beta_1(s)) - \hat{\mathbb{M}}_{nM}(\beta_2(s))| \\
&\leqslant \mathrm{E}\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\tilde{K}_h(s_m - s)[\widehat{S}_{\text{eff}}(\beta_1(s); X_i, Y_i(s)) - \widehat{S}_{\text{eff}}(\beta_2(s); X_i, Y_i(s))]\right\| \\
&= \mathrm{E}\left\|(\beta_1(s) - \beta_2(s))\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\tilde{K}_h(s_m - s)\int_0^1 \partial\widehat{S}_{\text{eff}}(\beta_1(s) + t(\beta_2(s) - \beta_1(s)))dt\right\| \\
&\leqslant C_1\|\beta_1(s) - \beta_2(s)\|,
\end{aligned}$$

where $\|\cdot\|_\psi$ is Orlicz norm, ∂ is the first derivative, and C_1 is a positive constant. Then applying

Theorem 2.2.4 in Van der Vaart and Wellner (1996) we can obtain that for any δ and ξ

$$\begin{aligned}
&E\left(\sup_{d(\beta_1(s),\beta_2(s))<\delta}|\hat{\mathbb{M}}_{nM}(\beta_1(s))-\hat{\mathbb{M}}_{nM}(\beta_2(s))|\right)\\
&\leqslant \left\|\sup_{d(\beta_1(s),\beta_2(s))<\delta}|\hat{\mathbb{M}}_{nM}(\beta_1(s))-\hat{\mathbb{M}}_{nM}(\beta_2(s))|\right\|_\psi\\
&\leqslant C_2\left[\int_0^\xi \psi^{-1}(D(\omega,\Theta,\|\cdot\|))d\omega+\delta\psi^{-1}(D^2(\xi,\Theta,\|\cdot\|))\right]\\
&\leqslant C_2\left[\int_0^\xi N^{\frac{1}{2p}}(\omega/2,\Theta,\|\cdot\|)d\omega+\delta N^{\frac{1}{p}}(\xi/2,\Theta,\|\cdot\|)\right),
\end{aligned}$$

where C_2 is a constant, $D(\cdot,\cdot,\cdot)$ and $N(\cdot,\cdot,\cdot)$ are the packing and bracketing number, respectively. Since $\beta(s)\in[-1,1]^p$ by definition, we have

$$N(\omega/2,\Theta,\|\cdot\|)\leqslant C_3\omega^{-p},$$

where C_3 is a constant. Consequently

$$E\left(\sup_{d(\beta_1(s),\beta_2(s))<\delta}|\hat{\mathbb{M}}_{nM}(\beta_1(s))-\hat{\mathbb{M}}_{nM}(\beta_2(s))|\right)\leqslant C_4\left[\int_0^\xi \omega^{-1/2}d\omega+\delta\xi^{-1}\right],$$

where C_4 is a constant. For any $\epsilon>0$, Markov's inequality gives

$$p\left(\sup_{d(\beta_1(s),\beta_2(s))<\delta}|\hat{\mathbb{M}}_{nM}(\beta_1(s))-\hat{\mathbb{M}}_{nM}(\beta_2(s))|>\epsilon\right)\leqslant\frac{C_4}{\epsilon}\left[2\xi^{1/2}+\delta\xi^{-1}\right].$$

By choosing first ξ and next δ, one can make the right hand side sufficient small, this verifies the asymototic equicontinuty of $\hat{\mathbb{M}}_{nM}(\cdot)$.

Secondly, following Wu and Zhang (2002), and employing Lemma 1 we can obtain

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\sum_{m=1}^M\tilde{K}_h(s_m-s)\widehat{S}_{\text{eff}}(\beta(s);X_i,Y_i(s_m))-\frac{1}{n}\sum_{i=1}^n\widehat{S}_{\text{eff}}(\beta(s);X_i,Y_i(s))\\
&\xrightarrow{p}\frac{\sum_{m=1}^M K_h(s_m-s)[E(S_{\text{eff}}(\beta(s);X_i,Y_i(s_m))-E(S_{\text{eff}}(\beta(s);X_i,Y_i(s))]}{\sum_{m=1}^M K_h(s_m-s)}\\
&=\frac{Mh\int K(\mu)[E(S_{\text{eff}}(\beta(s);X_i,Y_i(s+h\mu))-E(S_{\text{eff}}(\beta(s);X_i,Y_i(s))]\pi(s+h\mu)d\mu}{Mh\int K(\mu)\pi(s+h\mu)d\mu[1+O_p(Mh^{-1/2})]}\\
&\quad\times[1+O_p((Mh)^{-1/2}\sqrt{|\log h|})]\\
&=\frac{\int K(\mu)[\partial E(S_{\text{eff}}(s))h\mu+0.5\partial^2E(S_{\text{eff}}(s))h^2\mu^2+o(h^2)][\pi(s)+\dot{\pi}(s)h\mu+o(h)]d\mu}{\int K(\mu)[\pi(s)+\dot{\pi}(s)h\mu+o(h)]d\mu}[1+o_p(1)]\\
&=\frac{[0.5\partial^2E(S_{\text{eff}}(s))\pi(s)+\partial E(S_{\text{eff}}(s))\dot{\pi}(s)+o(h)]h^2\mu_2(K)}{\pi(s)+o(h)}[1+o_p(1)]\\
&=0.5h^2[\partial^2E(S_{\text{eff}}(s))+2\partial E(S_{\text{eff}}(s))\dot{\pi}(s)/\pi(s)]\mu_2(K)[1+o_p(1)],
\end{aligned}\tag{6}$$

where ∂, ∂^2 are the first and second derivative, and the term $[1 + o_p(1)]$ appears here because of the randomness of S_M. Consequently, for any $\beta(s) \in \Theta$

$$
\begin{aligned}
&|\hat{\mathbb{M}}_{nM}(\beta(s)) - \mathbb{M}(\beta(s))| \\
&\leqslant \left| \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{M} \tilde{K}_h(s_m - s) \widehat{S}_{\text{eff}}(\beta(s); X_i, Y_i(s_m)) - E(S_{\text{eff}}(\beta(s); X_i, Y_i(s))) \right| \\
&= \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{S}_{\text{eff}}(\beta(s); X_i, Y_i(s)) - E(S_{\text{eff}}(\beta(s); X_i, Y_i(s))) \right| + O_p(h^2) \\
&\leqslant \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{S}_{\text{eff}}(\beta(s); X_i, Y_i(s)) - \frac{1}{n} \sum_{i=1}^{n} S_{\text{eff}}(\beta(s); X_i, Y_i(s)) \right| \\
&\quad + \left| \frac{1}{n} \sum_{i=1}^{n} S_{\text{eff}}(\beta(s); X_i, Y_i(s)) - E(S_{\text{eff}}(\beta(s); X_i, Y_i(s))) \right| + O_p(h^2). \qquad (7)
\end{aligned}
$$

In (7), the first term tends to 0 in probability due to the consistency of $\widehat{E}\{X|X^T\beta(\cdot)\}$ and $\widehat{g}(X^T\beta_0(s_m))$ (Ma and Zhu, 2014) and the continuous mapping theorem under Assumption (C4), the convergence of the second term follows from the law of large numbers, and $O_p(h^2) = o_p(1)$ under Assumption (C9). These complete the marginal convergence of $\hat{\mathbb{M}}_{nM}(\cdot)$.

Finally, applying Lemma 2.8 in Newey and McFadden (1994), under Assumption (C4) the uniform convergence of $\hat{\mathbb{M}}_{nM}(\cdot)$ follows from its asymototic equicontinuty and marginal convergence.

Step 3: To prove that

$$
\sup_{\beta(s_m): d(\hat{\beta}(s_m), \beta_0(s_m)) > \epsilon} \sup_{s_m \in \mathcal{S}_M} \mathbb{M}(\hat{\beta}(s_m)) < \sup_{s_m \in \mathcal{S}_M} \mathbb{M}(\beta_0(s)). \qquad (8)
$$

This can be done by further using Assumptions (C3) and (C4).

Step 4: To prove that

$$
P\left(\sup_{s_m \in \mathcal{S}_M} d(\hat{\beta}(s_m), \beta_0(s_m)) > \epsilon \right) \to 0.
$$

The proof follows the same arguments used in Theorem 5.7 of Van der Vaart (1998). Since (5) implies $\sup_{s_m \in \mathcal{S}_M} |\hat{\mathbb{M}}_{nM}(\beta_0(s_m)) - \mathbb{M}(\beta_0(s_m))| \xrightarrow{p} 0$, so that under Assumption (C3), we have

$\hat{\mathbb{M}}_{nM}(\beta_0(s_m)) > \mathbb{M}(\hat{\beta}(s_m))$ for each $s_m \in \mathcal{S}_M$. It yields that

$$\begin{aligned}
&\sup_{s_m \in \mathcal{S}_M} [\mathbb{M}(\beta_0(s_m)) - \mathbb{M}(\hat{\beta}(s_m))] \\
&\leqslant \sup_{s_m \in \mathcal{S}_M} [\mathbb{M}(\beta_0(s_m)) - \hat{\mathbb{M}}_{nM}(\beta_0(s_m))] + \sup_{s_m \in \mathcal{S}_M} [\hat{\mathbb{M}}_{nM}(\beta_0(s_m)) - \mathbb{M}(\hat{\beta}(s_m))] \\
&\leqslant o_p(1) + \sup_{s_m \in \mathcal{S}_M} [\hat{\mathbb{M}}_{nM}(\hat{\beta}(s_m)) - \mathbb{M}(\hat{\beta}(s_m))] \\
&\leqslant o_p(1) + \sup_{\beta(s) \in \Theta} |\hat{\mathbb{M}}_{nM}(\beta(s)) - \mathbb{M}(\beta(s))| \\
&\xrightarrow{p} 0. \qquad (9)
\end{aligned}$$

From (8) we know that for any $\epsilon > 0$, there exists a positive constant δ that is only dependent on ϵ such that

$$\sup_{s_m \in \mathcal{S}_M} \mathbb{M}(\hat{\beta}(s_m)) \leqslant \sup_{s_m \in \mathcal{S}_M} \mathbb{M}(\beta_0(s)) - \delta(\epsilon)$$

for every $\hat{\beta}(s_m)$ when $\sup_{s_m \in \mathcal{S}_M} d(\hat{\beta}(s_m), \beta_0(s_m)) > \epsilon$. Consequently, by (9) we have

$$\begin{aligned}
P\left(\sup_{s_m \in \mathcal{S}_M} d(\hat{\beta}(s_m), \beta_0(s_m)) > \epsilon\right) &\leqslant P\Big(\sup_{s_m \in \mathcal{S}_M} \mathbb{M}(\hat{\beta}(s_m)) < \sup_{s_m \in \mathcal{S}_M} \mathbb{M}(\beta_0(s_m)) - \delta(\epsilon)\Big) \\
&\leqslant P\Big(\sup_{s_m \in \mathcal{S}_M} [\mathbb{M}(\beta_0(s_m)) - \mathbb{M}(\hat{\beta}(s_m))] > \delta(\epsilon)\Big) \\
&\to 0.
\end{aligned}$$

□

LEMMA 3: *Under assumption (C2) and (C5)-(C7),*

$$\sup_{s \in \mathcal{S}} |\Delta(s,h)| = o_p(1).$$

Proof of Lemma 3: By Lemma 1 with $r = 0$, we have

$$\begin{aligned}
\sup_{s \in \mathcal{S}_0} \left| \frac{1}{M} \sum_{m=1}^{M} K_h(s_m - s) - K_0(s,h) \right| &= \sup_{s \in \mathcal{S}_0} \left| \int K_h(t-s) d[\Pi_M(t) - \Pi(t)] \right| \\
&= O_p((Mh)^{-1/2}),
\end{aligned}$$

so that $\Delta(s,h)$ can be approximated by

$$
\begin{aligned}
\Delta(s,h_1) &= \frac{1}{K_0(s,h)+O_p((Mh)^{-1/2})}\left[\frac{1}{M}\sum_{m=1}^{M}K_h(s_m-s)\eta(s_m)\right.\\
&\quad\left.-\int K_h(t-s)\eta(t)d\Pi(t)\right]\\
&= \frac{1+O_p((Mh)^{-1/2})}{K_0(s,h)}\left\{\frac{1}{M}\sum_{m=1}^{M}K_h(s_m-s)[\eta(s_m)-\eta(s)]\right.\\
&\quad+\left[\frac{1}{M}\sum_{m=1}^{M}K_h(s_m-s)-\int K_h(t-s)d\Pi(t)\right]\eta(s)\\
&\quad\left.+\int K_h(t-s)[\eta(s)-\eta(t)]d\Pi(t)\right\}\\
&= \frac{1+o_p(1)}{K_0(s,h)}[(\mathrm{I})+(\mathrm{II})+(\mathrm{III})].
\end{aligned}
$$

Due to assumption (C5), $\eta(s)$ converges weakly to a Gaussian process, it follows from Donsker Theorem (Van der Vaart and Wellner, 1996) that

$$
\sup_{s\in\mathcal{S}}\|\eta(s)\|_2 = O_p(1).
$$

Thus we can examine the three terms as follows:

$$
\begin{aligned}
\frac{(\mathrm{I})}{K_0(s,h)} &\leqslant \frac{1}{K_0(s,h)}\frac{1}{M}\sum_{m=1}^{M}K_h(s_m-s)|\eta(s_m)-\eta(s)|\\
&\leqslant \sup_{|s'-s|\leqslant h}|\eta(s')-\eta(s)|\sup_{s\in\mathcal{S}}\frac{1}{MK_0(s,h)}\sum_{m=1}^{M}K_h(s_m-s)\\
&= o_p(1)\times\frac{K_0(s,h_1)+O_p((Mh_1)^{-1/2})}{K_0(s,h)}\\
&= o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
\frac{(\mathrm{II})}{K_0(s,h)} &\leqslant \frac{1}{K_0(s,h)}\sup_{s\in\mathcal{S}}|\eta(s)|\sup_{s\in\mathcal{S}}\left|\int K_h(t-s)d[\Pi_M(t)-\Pi(t)]\right|\\
&= O_p(h)\times O_p(1)\times O_p((Mh)^{-1/2})\\
&= o_p(1),
\end{aligned}
$$

$$\begin{aligned}\frac{\text{(III)}}{K_0(s,h)} &\leqslant \frac{1}{K_0(s,h)}\int K_h(t-s)|\eta(s)-\eta(t)|d\Pi(t) \\ &\leqslant \sup_{|s'-s|\leqslant h}|\eta(s')-\eta(s)|\sup_{s\in\mathcal{S}}\frac{1}{K_0(s,h)}\int K_h(t-s)d\Pi(t) \\ &= o_p(1)\times 1 \\ &= o_p(1).\end{aligned}$$

This finishes the proof of Lemma 3. □

Define $\overline{\varepsilon}_i(s) = \sum_{m=1}^{M}\tilde{K}_M(s_m - s, h_2)\varepsilon_i(s_m)$.

LEMMA 4: *Under Assumptions (C1)-(C11),*

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\overline{\varepsilon}_i(s)\eta_i(t)| = O_p(n^{-1/2}(\log n)^{1/2}),$$

and

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\overline{\varepsilon}_i(s)\overline{\varepsilon}_i(t)| = O((Mh_2)^{-1} + (\log n/n)^{1/2}) = o_p(1).$$

Proof of Lemma 4: The proof follows from a similar argument to Lemmas 6 and 7 of Zhu et al. (2012), so we just omit it here. □

Proof of Theorem 1

Theorem 1 (i) directly follows from (18) in the parent paper by noting that $D(\mathbf{w}(s))$ can be written as

$$D(\mathbf{w}(s)) = \frac{\mathbf{w}(s)^T(\Sigma_{\eta*,M} + \Lambda_{\epsilon*,M})\mathbf{w}(s)}{\mathbf{w}(s)^T\mathbf{1}_M\mathbf{1}_M^T\mathbf{w}(s)}.$$

We define $\Psi_{*,M} = [\psi_{1*,M}\cdots\psi_{M*,M}]$ and $\Lambda_{*,M} = \text{diag}(\lambda_{1*,M},\cdots,\lambda_{M*,M})$. We have

$$\begin{aligned}\Sigma_{\epsilon*,M}^{-1} &= \Lambda_{\epsilon*,M}^{-1/2}(\Lambda_{\epsilon*,M}^{-1/2}\Sigma_{\eta*,M}\Lambda_{\epsilon*,M}^{-1/2} + \mathbf{I}_M)^{-1}\Lambda_{\epsilon*,M}^{-1/2} \\ &= \Lambda_{\epsilon*,M}^{-1/2}(\Psi_{*,M}\Lambda_{*,M}\Psi_{*,M}^T + \mathbf{I}_M)^{-1}\Lambda_{\epsilon*,M}^{-1/2} \\ &= \Lambda_{\epsilon*,M}^{-1/2}\Psi_{*,M}(\Lambda_{*,M} + \mathbf{I}_M)^{-1}\Psi_{*,M}^T\Lambda_{\epsilon*,M}^{-1/2} \\ &= \Lambda_{\epsilon*,M}^{-1/2}\Psi_{*,M}\{(\Lambda_{*,M} + \mathbf{I}_M)^{-1} - \mathbf{I}_M\}\Psi_{*,M}^T\Lambda_{\epsilon*,M}^{-1/2} + \Lambda_{\epsilon*,M}^{-1}.\end{aligned}$$

By noting that $1/(\lambda_{m*,M}+1)-1=-\lambda_{m*,M}/(\lambda_{m*,M}+1)$, we can prove (2). □

Proof of Theorem 2

Proof of Theorem 2 (i): By the uniform consistency in Lemma 2, we can follow the arguments used in Theorem 3 of Ma and Zhu (2014) to obtain

$$\begin{aligned}0 &= n^{-1/2}\sum_{i=1}^{n}\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\varepsilon^*(s_m)[X_i-E\{X_i|X_i^T\beta(s)\}]\dot{g}(X_i^T\beta(s))\\ &\quad+n^{-1/2}\sum_{i=1}^{n}\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\varepsilon^*(s_m)[X_i-E\{X_i|X_i^T\beta(s)\}]X_i^T[\hat{\beta}(s)-\beta_0(s)]\ddot{g}(X_i^T\beta(s))\\ &\quad+o_p(1), \qquad (10)\end{aligned}$$

then by (10) we have

$$\begin{aligned}&\sqrt{n}[\hat{\beta}(s)-\beta_0(s)]\\ &=\left(-\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\varepsilon^*(s_m)[X_i-E\{X_i|X_i^T\beta(s)\}]X_i^T\ddot{g}(X_i^T\beta(s))\right)^{-1}\\ &\quad\times\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\varepsilon^*(s_m)[X_i-E\{X_i|X_i^T\beta(s)\}]\dot{g}(X_i^T\beta(s))+o_p(1)\right). \qquad (11)\end{aligned}$$

For notational simplicity, we denote

$$\begin{aligned}A_n(s)&=\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\varepsilon^*(s_m)[X_i-E\{X_i|X_i^T\beta(s)\}]X_i^T\ddot{g}(X_i^T\beta(s))\\ B_i(s)&=[X_i-E\{X_i|X_i^T\beta(s)\}]\dot{g}(X_i^T\beta(s)),\end{aligned}$$

then (11) can be rewritten as

$$\begin{aligned}&\sqrt{n}[\hat{\beta}(s)-\beta_0(s)]\\ &=[A_n(s)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}B_i(s)\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\varepsilon^*(s_m)+o_p(1)\\ &=[A_n(s)+O_p(h^2)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}B_i(s)\left\{\sum_{m=1}^{M}\tilde{K}_h(s_m-s)[g(X_i^T\beta(s_m))-g(X_i^T\beta(s))]\right.\\ &\quad\left.\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\eta_i(s_m)+\sum_{m=1}^{M}\tilde{K}_h(s_m-s)\varepsilon_i(s_m)\right\}+o_p(1). \qquad (12)\end{aligned}$$

Again similarly as (6), the bias term is

$$\sum_{m=1}^{M} \tilde{K}_h(s_m - s)[g(X_i^T\beta(s_m)) - g(X_i^T\beta(s))]$$
$$= 0.5h^2[\ddot{g}(X_i^T\beta(s)) + 2\dot{g}(X_i^T\beta(s))\dot{\pi}(s)/\pi(s)]\mu_2(K)[1 + o_p(1)]. \tag{13}$$

Next we show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_i(s)\sum_{m=1}^{M} \tilde{K}_h(s_m - s)\eta_i(s_m) \Rightarrow G(s), \tag{14}$$

where $\Rightarrow$ denotes weak convergence of a sequence of stochastic process and $G(s)$ is a central Gaussian process indexed by $s \in \mathcal{S}$. It consists of two steps. In Step 1, it follows from the standard central limit theorem that for each $s \in \mathcal{S}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_i(s)\sum_{m=1}^{M} \tilde{K}_h(s_m - s)\eta_i(s_m) \xrightarrow{d} N\left(0, \frac{1}{n}\sum_{i=1}^{n} B_i(s)\Sigma_\eta(s,s)B_i(s)^T\right),$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Step 2 is to show the asymptotic tightness of $n^{-1/2}\sum_{i=1}^{n} B_i(s)\sum_{m=1}^{M} \tilde{K}_h(s_m - s)\eta_i(s_m)$. Noting that $\sum_{m=1}^{M} \tilde{K}_h(s_m - s)\eta(s_m)$ can be approximated by three terms as follows:

$$\begin{aligned}
&\sum_{m=1}^{M} \tilde{K}_h(s_m - s)\eta(s_m) \\
&= \Delta(s,h) + \frac{1}{K_0(s,h)}\int K_h(t-s)\eta(t)\pi(t)d(t) \\
&= \Delta(s,h) + \frac{1}{K_0(s,h)}V(s)\int K_h(t-s)\pi(t)d(t) \\
&\quad + \frac{1}{K_0(s,h)}\int K_h(t-s)[\eta(t) - \eta(s)]\pi(t)d(t) \\
&= \Delta(s,h) + \text{(I)} + \text{(II)}.
\end{aligned}$$

Lemma 3 implies $\Delta(s,h)$ converges to zero uniformly. (I) is asymptotic tight since

$$\begin{aligned}
\text{(I)} &= \frac{1}{K_0(s,h)}\eta(s)K_0(s,h) \\
&= \eta(s),
\end{aligned}$$

where $\eta(s)$ is a Gaussian process. And it follows that when $h \to 0$,

$$\begin{aligned}(\text{II}) &\leqslant \sup_{|t-s|\leqslant h_1} |\eta(t)-\eta(s)| \frac{1}{K_0(s,h)} \int K_h(t-s)\pi(t)d(t) \\ &= o_p(1) \times 1 \\ &= o_p(1).\end{aligned}$$

Combining them together we can obtain the tightness.

Moreover, it follows from Lemma 1 that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} B_i(s) \sum_{m=1}^{M} \tilde{K}_h(s_m - s)\varepsilon_i(s_m) = O_p(|\log(h)|^{1/2}(Mh)^{-1/2}) = o_p(1) \tag{15}$$

holds uniformly for all $s \in \mathcal{S}$. For weak convergence, when we focus on the asymptotic distribution of $\hat{\beta}(s)$, the remainder $o_p(1)$ can be removed. Consequently, taking (13), (14) and (15) into (12) we can obtain that with a probability tending to one,

$$\mathrm{E}[\hat{\beta}(s)] - \beta_0(s) = 0.5h^2 A_n(s)^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} B_i(s)[\ddot{g}(X_i^T\beta(s)) + 2\dot{g}(X_i^T\beta(s))\dot{\pi}(s)/\pi(s)]\mu_2(K) \right\},$$

and

$$\begin{aligned}&\mathrm{Cov}\{\sqrt{n}[\hat{\beta}(s) - \beta_0(s)], \sqrt{n}[\hat{\beta}(t) - \beta_0(t)]\} \\ &= A_n(s)^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} B_i(s)\Sigma_\eta(s,t)B_i(t)^T \right] A_n(t)^{-1}.\end{aligned}$$

This finishes the proof of Theorem 2 (i).

Proof of Theorem 2 (ii): First, by following the proof of Lemma 4 in Zhu et al. (2012), we can easily have

$$\begin{aligned}&\mathrm{E}[\hat{g}(X^T\hat{\beta}(s))] - g(X^T\beta(s)) = 0.5h_1^2\mu_2(K)\ddot{g}(X^T\beta(s))[1 + o_p(1)] \\ &\mathrm{Var}[\hat{g}(X^T\hat{\beta}(s))] = n^{-1}\Sigma_\eta(s,s)[1 + o_p(1)].\end{aligned} \tag{16}$$

Moreover, by (12) in the parent paper and Theorem 1, it is easy to see that

$$
\begin{aligned}
&\sqrt{n}\{\hat{g}(X^T\hat{\beta}(s)) - E[\hat{g}(X^T\hat{\beta}(s))]\} \\
=&\sqrt{n}[1\ 0]\hat{\Sigma}(X^T\hat{\beta}(s), h_1)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M} K_{h_1}(X_i^T\hat{\beta}(s_m) - X^T\hat{\beta}(s))\hat{Z}_{i,m,s}[\eta(s_m) + \varepsilon(s_m)] \\
=&\sqrt{n}[1\ 0]\Sigma(X^T\beta(s), h_1)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M} K_{h_1}(X_i^T\beta(s_m) - X^T\beta(s))Z_{i,m,s}[\eta(s_m) + \varepsilon(s_m)\} \\
&\times [1 + O_p(n^{-1/2} + h^2)].
\end{aligned}
$$

Then following the proof of Theorem 1 in Zhu et al. (2012), we can have that $\sqrt{n}[\hat{g}(X^T\hat{\beta}(s)) - g(X^T\beta(s)) - 0.5h_1^2\mu_2(K)\ddot{g}(X^T\beta(s))]$ converges to a central Gaussian process with covariance function $\Sigma_\eta(s,t)$. □

Proof of Theorem 3

Let $\tilde{K}_M(s,h) = \tilde{K}_M(s/h)/h$, where $\tilde{K}_M(s)$ is the empirical equivalent kernels for the first-order local polynomial kernel (Fan and Gijbels, 1996). Thus, we have

$$
\begin{aligned}
&\hat{\eta}_i(s) - \eta_i(s) \\
=&\sum_{m=1}^{M}\tilde{K}_M(s_m - s, h_2)[g(X_i^T\beta(s_m)) - \hat{g}(X_i^T\hat{\beta}(s_m)] \\
&+\sum_{m=1}^{M}\tilde{K}_M(s_m - s, h_2)[\eta_i(s_m) + \varepsilon_i(s_m) - \eta_i(s)]. \qquad (17)
\end{aligned}
$$

It follows from a Taylor's expansion that

$$
\sum_{m=1}^{M}\tilde{K}_M(s_m - s, h_2)[\eta_i(s_m) - \eta_i(s)] = 0.5\mu_2(K)\ddot{\eta}_i(s)h_2^2[1 + o_p(1)],
$$

and

$$
\begin{aligned}
&\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)\{g(X_i^T\beta(s_m)) - \mathrm{E}[\hat{g}(X_i^T\hat{\beta}(s_m))|\mathcal{S},\eta,X\} \\
=&\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)[0.5h_1^2\mu_2(K)\ddot{g}(X_i^T\beta(s_m)) + n^{-1}\sum_{i'=1}^{n}\eta_{i'}(s_m)] \\
&\times [1 + O_p(h_1 + n^{-1/2} + (Mh_1)^{-1/2})], \\
=&[0.5h_1^2\mu_2(K)\ddot{g}(X_i^T\beta(s_m)) + O_p(n^{-1/2})] \\
&\times [1 + O_p(h_1 + h_2 + n^{-1/2}) + O_p((Mh_2)^{-1/2} + (Mh_1)^{-1/2})],
\end{aligned}
$$

which leads to Bias$[\hat{\eta}_i(s)|\mathcal{S},\eta,X]$.

Furthermore, it can be shown that

$$
\begin{aligned}
&\hat{\eta}_i(s) - \mathrm{E}[\hat{\eta}_i(s)|\mathcal{S},X,\eta] \\
=&\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)\{\varepsilon_i(s_m) \\
&- [1\ 0]\hat{\Sigma}(X^T\beta(s), h_1)^{-1}\sum_{i',m'} K_{h_1}(X_{i'}^T\hat{\beta}(s_{m'}) - X_i^T\hat{\beta}(s_m))\hat{Z}_{i,m,s}\varepsilon_{i'}(s_{m'})\} \\
=&\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)\{\varepsilon_i(s_m) \\
&- [1\ 0]\Sigma(X^T\beta(s), h_1)^{-1}\sum_{i',m'} K_{h_1}(X_{i'}^T\beta(s_{m'}) - X_i^T\beta(s_m))Z_{i,m,s}\varepsilon_{i'}(s_{m'})\}[1 + o_p(1)].
\end{aligned}
$$

After some tedious calculations, we have

$$
\begin{aligned}
&\mathrm{Cov}(\hat{\eta}_i(s) - \eta_i(s), \hat{\eta}_i(t) - \eta_i(t)|\mathcal{S},\eta,X) \\
=&K^{\#}((s-t)/h_2)\sigma_\varepsilon(s)\pi(t)^{-1}(Mh_2)^{-1}[1 + o_p(1)] - (nMh_1)^{-1}\pi(s)^{-1}\pi(t)^{-1}O_p(1)
\end{aligned}
$$

Furthermore, for $i = 1, \cdots, n$, after dropping some higher order terms, we have

$$
\begin{aligned}
&\mathrm{E}\{[\hat{\eta}_i(s) - \eta_i(s)]^2|\mathcal{S},\eta,X\} \\
=&\{\mathrm{E}[\hat{\eta}_i(s) - \eta_i(s)|\mathcal{S},\eta,X]\}^2 + \mathrm{Var}[\hat{\eta}_i(s) - \eta_i(s)|\mathcal{S},\eta,X] \\
=&[0.5\mu_2(K)\ddot{\eta}_i(s)h_2^2 + 0.5\mu_2(K)\ddot{g}(X_i^T\beta(s_m))h_1^2 + n^{-1}\sum_{i'=1}^{n}\eta_{i'}(s_m)]^2[1 + o_p(1)] \\
&+ v_0(K)\pi(s)^{-1}\sigma_\varepsilon(s)(Mh_2)^{-1}[1 + o_p(1)] - (nMh_1)^{-1}\pi(s)^{-2}O_p(1)
\end{aligned}
$$

Combining the above results we obtain $\mathrm{Cov}(\hat{\eta}_i(s), \hat{\eta}_i(t)|\mathcal{S}, X, \eta,)$ and complete the proof of (i).

It follows from that

$$\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)\{g(X_i^T\beta(s_m)) - \mathrm{E}[\hat{g}(X_i^T\beta(s_m))|\mathcal{S}, X\}$$
$$= 0.5h_1^2\mu_2(K)\ddot{g}(X_i^T\beta(s_m))[1 + O_p(h_1 + n^{-1/2} + (Mh_1)^{-1/2})],$$

Furthermore, it can be shown that

$$\begin{aligned}
&\hat{\eta}_i(s) - \eta_i(s) - \mathrm{E}[\hat{\eta}_i(s)|\mathcal{S}, X] \\
=&\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)[\eta_i(s_m) + \varepsilon_i(s_m) - \eta_i(s)] - \sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2) \\
&\{[1\ 0]\hat{\Sigma}(X^T\beta(s), h_1)^{-1} \sum_{i',m'} K_{h_1}(X_{i'}^T\hat{\beta}(s_{m'}) - X_i^T\hat{\beta}(s_m))\hat{Z}_{i,m,s}[\eta_{i'}(s_{m'}) + \varepsilon_{i'}(s_{m'})]\} \\
=&\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)[\eta_i(s_m) + \varepsilon_i(s_m) - \eta_i(s)] - \sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2) \\
&\{[1\ 0]\Sigma(X^T\beta(s), h_1)^{-1} \sum_{i',m'} K_{h_1}(X_{i'}^T\beta(s_{m'}) - X_i^T\beta(s_m))Z_{i,m,s}[\eta_{i'}(s_{m'}) + \varepsilon_{i'}(s_{m'})]\}[1 + o_p(1)].
\end{aligned}$$

With tedious calculations, we have

$$\begin{aligned}
&\mathrm{Cov}(\hat{\eta}_i(s) - \eta_i(s), \hat{\eta}_i(t) - \eta_i(t)|\mathcal{S}, X) \\
=&K^{\#}((s-t)/h_2)\sigma_\varepsilon(s)\pi(t)^{-1}(Mh_2)^{-1}[1 + o_p(1)] - (nMh_1)^{-1}\pi(s)^{-1}\pi(t)^{-1}O_p(1) \\
&+ \{0.25\mu_2(K)^2h_2^4\Sigma_\eta^{(2,2)}(s,t) + n^{-1}\Sigma_\eta(s,t) \\
&- 0.5n^{-1}\mu_2(K)h_2^2[\Sigma_\eta^{(2,0)}(s,t)\pi(s)^{-1} + \Sigma_\eta^{(0,2)}(s,t)\pi(t)^{-1}]\}[1 + o_p(1)].
\end{aligned}$$

It follows from Lemma 1 and (16) that

$$\begin{aligned}
&\mathrm{E}\{[\hat{\eta}_i(s) - \eta_i(s)]^2|\mathcal{S}, X] \\
=&\{\mathrm{E}[\hat{\eta}_i(s) - \eta_i(s)|\mathcal{S}, X]\}^2 + \mathrm{Var}[\hat{\eta}_i(s) - \eta_i(s)|\mathcal{S}, X] \\
=&\{0.25\mu_2(K)^2h_1^4[\ddot{g}(X_i\beta_s)]^2 + 0.25\mu_2(K)^2h_2^4\Sigma_\eta^{(2,2)}(s,s) \\
&+ n^{-1}\Sigma_\eta(s,s) + v_0(K)\sigma_\varepsilon^2(s,s)\pi(s)^{-1}(Mh_2)^{-1}\}[1 + o_p(1)],
\end{aligned}$$

which leads to Theorem 3 (ii).

Furthermore, by noting that $\mathrm{E}\{[\hat{\eta}_i(s) - \eta_i(s)]^2|\mathcal{S}] = \mathrm{E}(\mathrm{E}\{[\hat{\eta}_i(s) - \eta_i(s)]^2|\mathcal{S}, X]|\mathcal{S})$, we can easily obtain Theorem 4 (iii) and (iv).

To show (v), we define

$$\begin{aligned}
\bar{\varepsilon}_i(s) &= \sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)\varepsilon_i(s_m), \\
\Delta\eta_i(s) &= \sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)[\eta_i(s_m) - \eta_i(s)], \\
\Delta g_i(s) &= \sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)[g(X_i^T\beta(s_m)) - \hat{g}(X_i^T\hat{\beta}(s_m))], \\
\Delta_i(s) &= \bar{\varepsilon}_i(s) + \Delta\eta_i(s) + \Delta g_i(s).
\end{aligned}$$

Recall from (17) that

$$\hat{\eta}_i(s) - \eta_i(s) = \Delta_i(s).$$

It follows from Lemma 2 and Taylor series expansion that

$$\sup_{s\in\mathcal{S}} |\bar{\varepsilon}_i(s)| = O_p(\sqrt{\frac{|\log(h_2)|}{Mh_2}}) \text{ and } \sup_{s\in\mathcal{S}} |\Delta\eta_i(s)| = O_p(1) \sup_{s\in\mathcal{S}} |\ddot{\eta}_i(s)| h_2^2.$$

By Theorem 2, the sequence $\sqrt{n}[\hat{g}(x^T\hat{\beta}(s)) - g(x^T\beta(s)) - 0.5h_1^2\mu_2(K)\ddot{g}(x^T\beta(s))]$ is asymptotically tight. Thus, we have

$$\begin{aligned}
&\Delta g_i(s) = -\sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)0.5h_1^2\mu_2(K)\ddot{g}(\cdot)[1 + o_p(1)] \\
&+ \sum_{m=1}^{M} \tilde{K}_M(s_m - s, h_2)[0.5h_1^2\mu_2(K)\ddot{g}(\cdot)[1 + o_p(1)] + g(X_i^T\beta(s_m)) - \hat{g}(X_i^T\hat{\beta}(s_m))], \\
&\sup_{s\in\mathcal{S}} |\Delta g_i(s)| = O_p(h_1^2) + O_p(n^{-1/2}).
\end{aligned}$$

Combining these results, we have

$$\sup_{s\in\mathcal{S}} |\hat{\eta}_i(s) - \eta_i(s)| = O_p(|\log(h_2)|^{1/2}(Mh_2)^{-1/2} + h_1^2 + h_2^2 + n^{-1/2}).$$

□

Proof of Theorem 4

Recall that $\hat{\eta}_i(s) = \eta_i(s) + \Delta_i(s)$, we have

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\hat{\eta}_i(s)\hat{\eta}_i(t) =& n^{-1}\sum_{i=1}^{n}\Delta_i(s)\Delta_i(t) + n^{-1}\sum_{i=1}^{n}\eta_i(s)\Delta_i(t) \\ &+ n^{-1}\sum_{i=1}^{n}\Delta_i(s)\eta_i(t) + n^{-1}\sum_{i=1}^{n}\eta_i(s)\eta_i(t). \end{aligned} \tag{18}$$

This proof consists of two steps. The first step is to show that the first three terms on the right hand side of (18) converge to zero uniformly for all $(s,t) \in \mathcal{S}^2$ in probability. The second step is to show the uniform convergence of $n^{-1}\sum_{i=1}^{n}\eta_i(s)\eta_i(t)$ to $\Sigma_\eta(s,t)$ over $(s,t) \in \mathcal{S}^2$ in probability.

We first show that

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\Delta_i(s)\eta_i(t)| = O_p(n^{-1/2} + h_1^2 + h_2^2 + (\log n/n)^{1/2}). \tag{19}$$

Since

$$\sum_{i=1}^{n}\Delta_i(s)\eta_i(t) \leqslant |\sum_{i=1}^{n}\bar{\varepsilon}_i(s)\eta_i(t)| + |\sum_{i=1}^{n}\Delta\eta_i(s)\eta_i(t)| + |\sum_{i=1}^{n}\Delta g_i(s)\eta_i(t)|, \tag{20}$$

it is sufficient to focus on the three terms on the right-hand side of (20). Since

$$|\Delta g_i(s)\eta_i(t)| \leqslant \sup_{s\in\mathcal{S}}|\Delta g_i(s)|\sup_{t\in\mathcal{S}}|\eta_i(t)|,$$

we have

$$n^{-1}|\sum_{i=1}^{n}\Delta g_i(s)\eta_i(t)| \leqslant n^{-1}\sum_{i=1}^{n}\sup_{s,t\in\mathcal{S}}|\Delta g_i(s)\eta_i(t)| = O_p(h_1^2 + n^{-1/2}) = o_p(1).$$

Similarly, we have

$$n^{-1}|\sum_{i=1}^{n}\Delta\eta_i(s)\eta_i(t)| \leqslant n^{-1}\sum_{i=1}^{n}\sup_{s,t\in\mathcal{S}}|\Delta\eta_i(s)\eta_i(t)| = O_p(h_2^2) = o_p(1).$$

It follows from Lemma 4 that $\sup_{(s,t)} n^{-1}\{|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)\eta_i(t)| = O((\log n/n)^{1/2})$. Similarly, we can show that $\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\Delta_i(t)\eta_i(s)| = O_p(n^{-1/2} + h_1^2 + h_2^2 + (\log n/n)^{1/2})$.

We next show that

$$\sup_{(s,t)}|n^{-1}\sum_{i=1}^{n}[\eta_i(s)\eta_i(t) - \Sigma_\eta(s,t)]| = O_p(n^{-1/2}). \tag{21}$$

This can be argued by noting that

$$\begin{aligned} &|\eta_i(s_1)\eta_i(t_1) - \eta_i(s_2)\eta_i(t_2)| \\ \leqslant& 2(|s_1 - s_2| + |t_1 - t_2|)\sup_{s\in\mathcal{S}}|\dot{\eta}_i(s)|\sup_{s\in\mathcal{S}}|\eta_i(s)| \end{aligned}$$

holds for any (s_1, t_1) and (s_2, t_2). Therefore the functional class $\{\eta(u)\eta(v) : (u, v) \in \mathcal{S}^2\}$ is a Vapnik and Cervonenkis (VC) class (Van der Vaart, 1998). Thus, it yields that (21) is true.

Finally, we can show that

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n} \Delta_i(s)\Delta_i(t)| = O_p((Mh_2)^{-1} + (\log n/n)^{1/2} + h_1^4 + h_2^4). \tag{22}$$

With some calculations, for a positive constant C_1, we have

$$\begin{aligned}
&|\sum_{i=1}^{n} \Delta_i(s)\Delta_i(t)| \\
\leqslant &C_1 \sup_{(s,t)}[|\sum_{i=1}^{n} \overline{\varepsilon}_i(s)\overline{\varepsilon}_i(t)| + |\sum_{i=1}^{n} \overline{\varepsilon}_i(s)\Delta\eta_i(t)| + |\sum_{i=1}^{n} \Delta\eta_i(t)\Delta g_i(s)| \\
&+ |\sum_{i=1}^{n} \overline{\varepsilon}_i(s)\Delta g_i(t)| + |\sum_{i=1}^{n} \Delta g_i(s)\Delta g_i(t)| + |\sum_{i=1}^{n} \Delta\eta_i(s)\Delta\eta_i(t)|.
\end{aligned}$$

It follows from Lemma 4 that

$$\begin{aligned}
&\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n} \overline{\varepsilon}_i(s)\overline{\varepsilon}_i(t)| = O_p((Mh_2)^{-1} + (\log n/n)^{1/2}), \\
&\sup_{(s,t)} n^{-1}[|\sum_{i=1}^{n} \overline{\varepsilon}_i(s)\Delta\eta_i(t)| + |\sum_{i=1}^{n} \Delta\eta_i(t)\Delta g_i(s)| + |\sum_{i=1}^{n} \overline{\varepsilon}_i(s)\Delta g_i(t)| = O_p((\log n/n)^{1/2}).
\end{aligned}$$

Since $\sup_{s\in\mathcal{S}} |\Delta\eta_i(s)| = C_2 \sup_{s\in\mathcal{S}} |\ddot{\eta}_i(s)|h_2^2$, we have $\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n} \Delta\eta_i(s)\Delta\eta_i(t)| = O(h_2^4)$. Furthermore, we have

$$n^{-1}|\sum_{i=1}^{n} \Delta g_i(s)\Delta g_i(t)| = O_p(n^{-1} + h_1^4)$$

Thus, combining (19)-(22)) leads to Theorem 4 (i).

To prove Theorem 4 (ii), we may follow the same arguments in Lemma 6 of Li and Hsing (2010). For completion, we highlight several key steps below. We define

$$\Delta\psi(s) = \int_0^1 [\hat{\Sigma}_\eta(s,t) - \Sigma_\eta(s,t)]\psi(t)dt. \tag{23}$$

Following Hall et al. (2006) and the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\{\int_0^1 [\hat{\psi}(s) - \psi(s)]^2 ds\}^{1/2} \\
\leqslant & C_2\{[\int_0^1 \Delta\psi(s)^2 ds]^{1/2} + \int_0^1\int_0^1 [\hat{\Sigma}_\eta(s,t) - \Sigma_\eta(s,t)]^2 dsdt\} \\
\leqslant & C_2\{\int_0^1\int_0^1 [\hat{\Sigma}_\eta(s,t) - \Sigma_\eta(s,t)]^2 dsdt\}^{1/2}\{\int_0^1 [\psi(t)]^2 dt\}^{1/2} \\
& + \int_0^1\int_0^1 [\hat{\Sigma}_\eta(s,t) - \Sigma_\eta(s,t)]^2 dsdt \\
\leqslant & C_3 \sup_{(s,t)\in\mathcal{S}^2} |\hat{\Sigma}_\eta(s,t) - \Sigma_\eta(s,t)|,
\end{aligned}$$

which yields Theorem 4 (ii.a).

Using (4.9) in Hall et al. (2006), we have

$$\begin{aligned}
&|\hat{\lambda} - \lambda| \\
\leqslant & |\int_0^1\int_0^1 [\hat{\Sigma}_\eta - \Sigma_\eta](s,t)\psi(s)\psi(t) dsdt + O(\int_0^1 \Delta\psi(s)^2 ds) \\
\leqslant & C_4 \sup_{(s,t)\in\mathcal{S}^2} |\hat{\Sigma}_\eta(s,t) - \Sigma_\eta(s,t)|,
\end{aligned}$$

which yields Theorem 4 (ii.b). □

Web Appendix B: Figures

Typical estimated covariance function $\Sigma_\eta(\cdot,\cdot)$ and eigen-functions $\psi_i(\cdot), i = 1, 2$ are shown in Figure 1. The estimated covariance function and curves are quite close to their true values.

Figure 2 shows the Q-Q plots of $\hat{\eta}$ and $\hat{\varepsilon}$ at 3 randomly selected grid points. Their quantiles are quite close to the straight lines, which represent standard normal distribution. It verifies the Gaussian assumptions of individual curve variations and random errors.

Figure 3 presents the estimated seven varying coefficients their corresponding 95% simultaneous confidence bands. It reveals that MMSE, age, education level and AD status are the most important factors. Moreover, gender and handiness have little effects on FA.

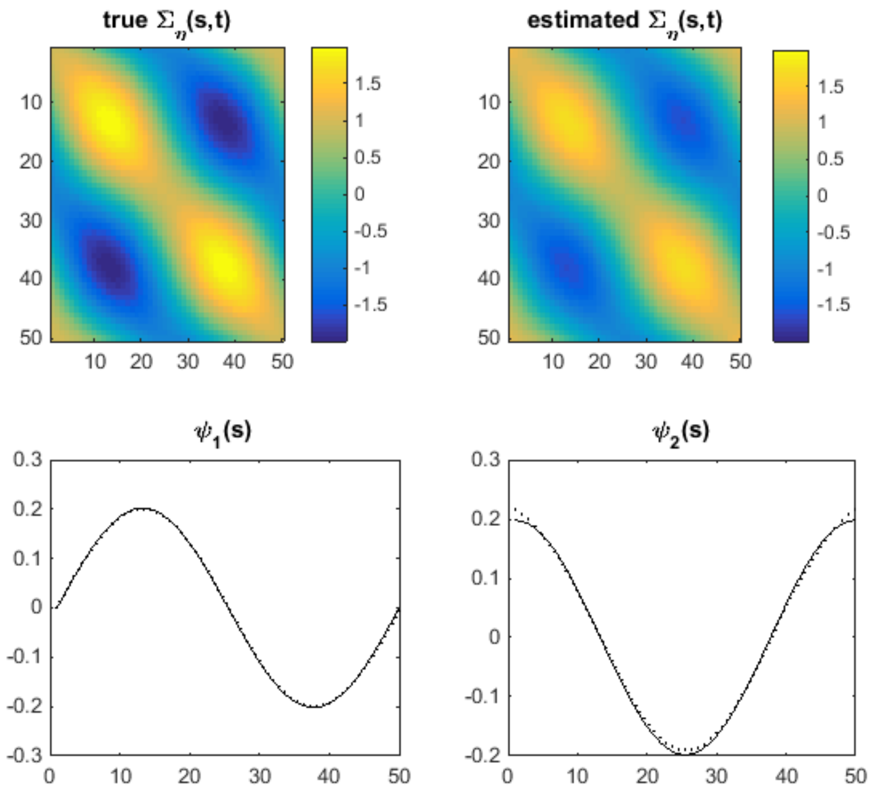

Figure 1. Simulation results for the first index function with $n = 200$: true and estimated Σ_η (top row), and the estimated eigenfunctions (bottom row). The solid lines are true functions and the broken lines are estimated functions.

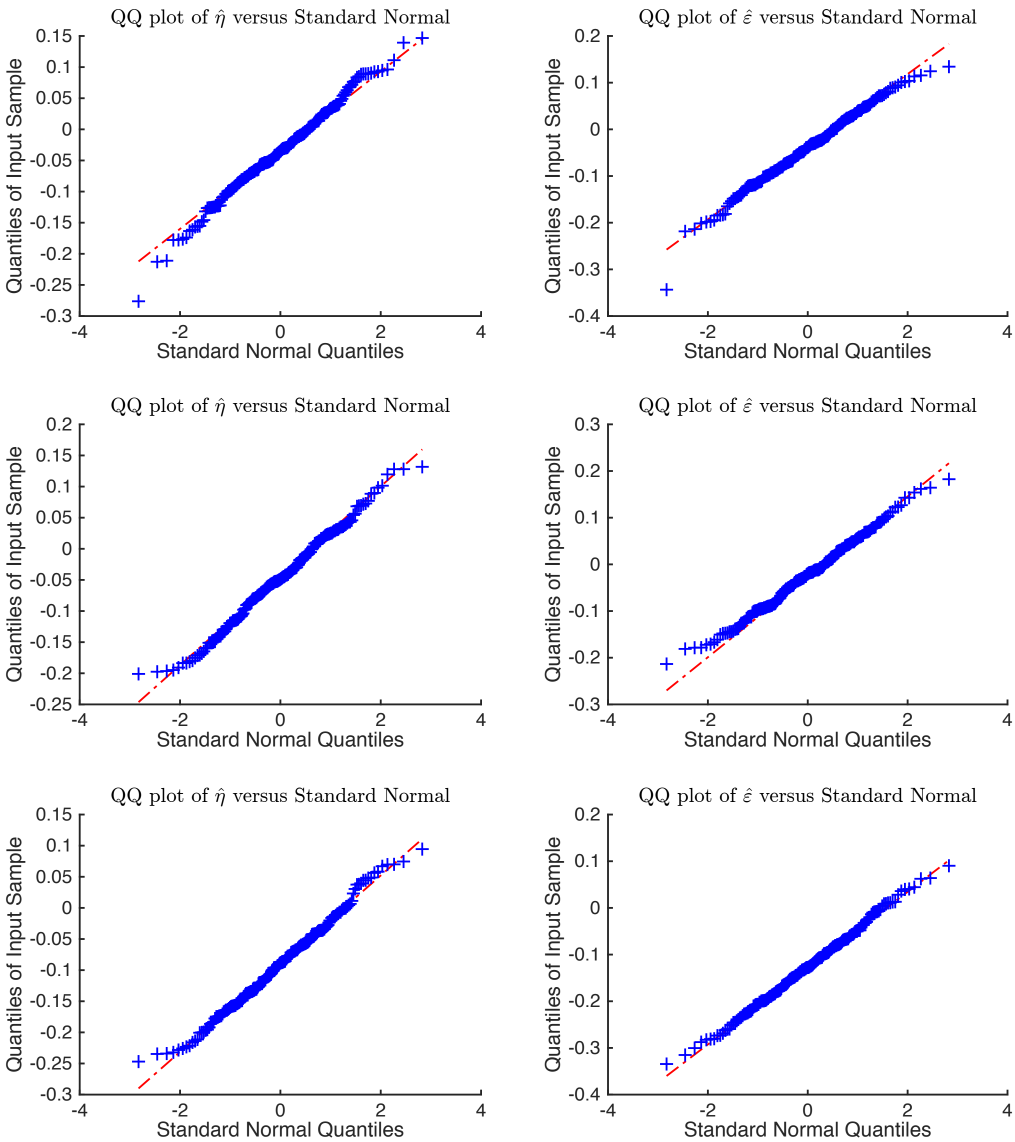

Figure 2. ADNI data analysis: the qq plots of $\hat{\eta}$ (left column) and $\hat{\varepsilon}$ (right column) at grid points 20 (top row), 40 (middle row) and 60 (bottom row).

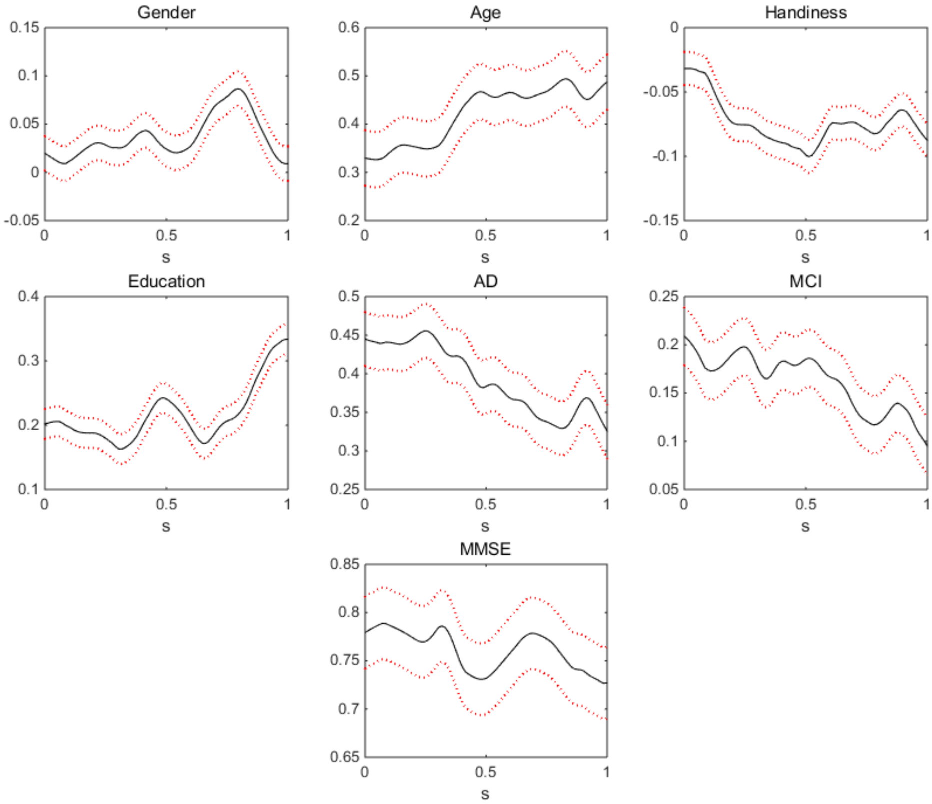

Figure 3. ADNI data analysis: the seven estimated varying coefficients. The black solid lines are estimated coefficients and the red broken lines are their corresponding 95% simultaneous confidence bands.

References

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: Monographs on statistics and applied probability 66*, volume 66. CRC Press.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics* pages 1493–1517.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* pages 3321–3351.

Li, Y., Hsing, T., et al. (2010). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *The Annals of Statistics* **38,** 3028–3062.

Ma, Y. and Zhu, L. (2014). On estimation efficiency of the central mean subspace. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **76,** 885–901.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics* **4,** 2111–2245.

Van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Wu, H. and Zhang, J.-T. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association* **97,** 883–897.

Zhu, H., Li, R., and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *Annals of Statistics* **40,** 2634–2666.